Structure
Resource

Tracing Beta Strands Using StrandTwister from Cryo-EM Density Maps at Medium Resolutions

Dong Si[1] and Jing He[1,*]
[1]Department of Computer Science, Old Dominion University, Norfolk, VA 23529, USA
*Correspondence: jhe@cs.odu.edu
http://dx.doi.org/10.1016/j.str.2014.08.017

SUMMARY

Major secondary structure elements such as α helices and β sheets can be computationally detected from cryoelectron microscopy (cryo-EM) density maps with medium resolutions of 5–10 Å. However, a critical piece of information for modeling atomic structures is missing, because there are no tools to detect β strands from cryo-EM maps at medium resolutions. We propose a method, StrandTwister, to detect the traces of β strands through the analysis of twist, an intrinsic nature of a β sheet. StrandTwister has been tested using 100 β sheets simulated at 10 Å resolution and 39 β sheets computationally detected from cryo-EM density maps at 4.4–7.4 Å resolutions. Although experimentally derived cryo-EM maps contain errors, StrandTwister's best detections over 39 cases were able to detect 81.87% of the β strands, with an overall 1.66 Å two-way distance between the detected and observed β traces. StrandTwister appears to detect the traces of β strands on major β sheets quite accurately, particularly at the central area of a β sheet.

INTRODUCTION

Cryoelectron microscopy (cryo-EM) has become a major experimental technique to study the structures of large protein complexes (Chiu et al., 2005; Hryc et al., 2011). It is a structure determination technique complementary to X-ray crystallography and nuclear magnetic resonance. A number of large molecular complexes, such as ribosome and viruses, have been resolved to near atomic resolutions (2–5 Å) (Anger et al., 2013; Jiang et al., 2008; Yu et al., 2008; Zhang et al., 2013). Many more have reached medium resolutions (5–10 Å) (Esquivel-Rodríguez and Kihara, 2013; Lawson et al., 2011). At medium resolutions, molecular features are less resolved, and it is challenging to derive atomic structures from such density maps. In some special situations, particularly for small proteins with mostly α helices, direct modeling is possible to derive the backbone of a protein (Crowther et al., 1994). A major approach, though, is to start with a homologous model and adjust it through fitting (Beck et al., 2012; Hashem et al., 2013; Lasker et al., 2012; Zhang et al., 2010; Zhao et al., 2013). The initial model can be a homologous structure or a model built from a template structure (Arnold et al., 2006; Sali and Blundell, 1993; Topf et al., 2006). Fitting methods have evolved from earlier rigid fitting to flexible fitting (Chan et al., 2012; Rossmann, 2000; Schröder et al., 2007; Topf et al., 2008; Trabuco et al., 2008). Although fitting a homologous model has been fairly successful resolving structures from density maps at medium resolutions, it is still challenging to find suitable known structures as templates in many cases. De novo modeling aims to derive near atomic structures from electron density maps without a template structure. The connection between secondary structure elements, such as α helices and β sheets, is ambiguous at medium resolutions. However, the likely connections may be derived. Given the positions of α helices and β strands in a density map, one can match them with secondary structure sequence segments that can be predicted from the amino acid sequence to derive the overall topology of a protein chain (Al Nasr et al., 2011; Baker et al., 2011; Biswas et al., 2012; Lindert et al., 2012; Lu et al., 2008). Once the topology is determined, backbone and side chains can be constructed and evaluated using energy functions (Al Nasr et al., 2010; Lindert et al., 2012; Sun and He, 2009). Pathwalking derives a Cα trace directly from pseudoatoms extracted from the density map (Baker et al., 2012b). The main drawback of this method appears to be at β sheet regions where the β strands are not resolved.

The locations of secondary structures are critical in modeling atomic structure from density maps and as overall shape descriptors in identifying similar structures (Esquivel-Rodríguez and Kihara, 2013). Although it is not possible to distinguish the amino acid at medium resolutions, secondary structures such as α helices (red lines in Figure 1B) and β sheets (blue density voxels in Figure 1B) can be identified (Baker et al., 2007; Del Palu et al., 2006; Jiang et al., 2001; Kong and Ma, 2003; Rusu and Wriggers, 2012; Si et al., 2012; Yu and Bajaj, 2008). A β sheet contains multiple β strands. Although β sheets can be identified from cryo-EM density maps at 5 to 10 Å resolutions, it is almost impossible to detect the β strands of a β sheet. The spacing between two neighboring β strands is between 4.5 and 5 Å, and β strands are visible only when the resolution is higher than 4.7 Å (Baker et al., 2012a; Zhou, 2008). Without knowing the locations of β strands, the representation of protein is purely dependent on the relative locations of helices (Lasker et al., 2007). De novo modeling has been successful deriving the backbone from the density map of GroEL (4.2 Å resolution) (Ludtke et al., 2008) and gp10 (4.5 Å resolution) (Baker et al., 2013). However, there has not been an α/β structure that is resolved using de novo modeling from a density map at a medium resolution. One of the challenges is the inability to detect β strands from the density maps. In addition to the secondary structural elements, a

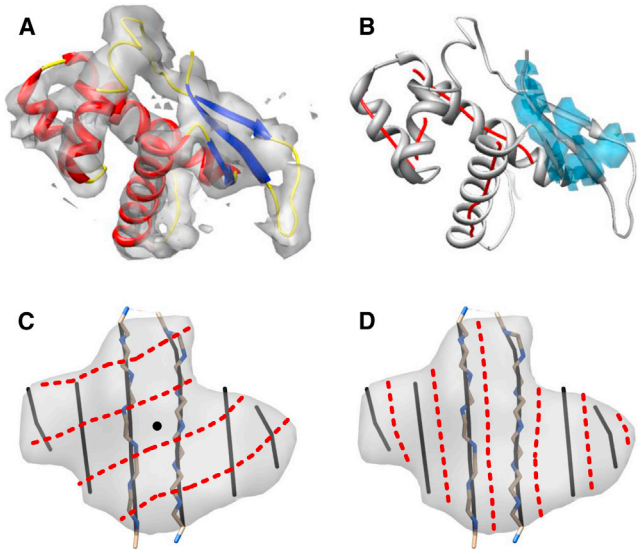

Figure 1. The Problem of β Strand Detection from Medium-Resolution Density Maps

(A) The density corresponding to chain R of 3FIN (PDB accession number) was extracted from cryo-EM density map EMD_5030 (6.4 Å resolution) and was superimposed with its corresponding PDB structure (ribbon).

(B–D) The detected helices (red lines) and β sheet region (surface view of blue voxels) in (B) using SSELearner (Si et al., 2012). Two possible sets of β traces (black solid lines and red dashed lines) may differ in orientation (C) and/or position (shift) (D). The backbone structure is superimposed on the observed β traces for the middle two β strands in (C) and (D).

skeleton that represents possible connections can be identified (Al Nasr et al., 2013; Ju et al., 2007), although ambiguous points exist in the skeleton.

A helix detected from the medium-resolution data is often represented as a line, referred here as an α trace that corresponds to the central axis of a helix (red lines in Figure 1B). We define a β trace (black line in Figures 1C and 1D) as the central line along a β strand. In particular, an observed β trace is the line interpolating all geometric centers of three consecutive Cα atoms on a β strand plus two Cα atoms at the end of the β strand. At medium resolutions, the Cα trace of a β strand is not resolved in the density map. The problem of detecting β strands from the density of a β sheet is to find the orientations (Figure 1C) and locations (Figure 1D) of β traces.

As more experimentally determined cryo-EM maps accumulate in the Electron Microscopy Data Bank (EMDB) (Henderson et al., 2012; Lawson et al., 2011), it is clear that such cryo-EM maps are more challenging than the density maps simulated at the same resolution. A simulated density map at 6 Å resolution often reveals the separation of β strands at a proper range of density thresholds (Figure 2A). Even at 8 Å resolution, a simulated density map partially reveals the separation of β strands (Figure 2B), and it has much better quality than an experimentally obtained cryo-EM map at a similar resolution. However, a simulated density map at 9 or 10 Å resolution is often challenging for visual detection (Figure 2C), and so is a cryo-EM map at 6.4 Å resolution (Figure 2D). The question we address is whether it is possible to derive β traces from cryo-EM maps at medium resolutions when no separation of β strands is detectable.

Currently there are no tools to derive the β traces from cryo-EM density maps at medium resolutions. Sheettracer was the first attempt to derive β traces (Kong et al., 2004). It uses deconvolution to enhance the separation of β strands, while filtering and clustering follow. However, this method was predominantly tested using simulated density maps with resolutions of 6 and 8 Å. For simulated density maps at such resolutions, the separation of β strands may be visible or partially visible (Figures 2A and 2B); however, this is not true for experimentally obtained cryo-EM maps (Figure 2D). Sheettracer may be suitable for density maps in which the separation of β strands is partially visible. Gorgon uses a semiautomated method allowing a user to determine the position of β strands, which is challenging to apply the cryo-EM maps at medium resolutions (Baker et al., 2010). Path-walking derives a backbone from cryo-EM maps at near atomic resolutions, such as 3 to 5 Å. It performs well in α helical domains but fails at β sheet regions at the medium resolutions (Baker et al., 2012b).

The detection of β strands from medium-resolution cryo-EM maps remains an open problem since the first attempt in 2004 (Kong et al., 2004). In addressing this challenge, we propose a method, StrandTwister, that does not rely on the separation of β strands and therefore is applicable to much lower resolutions. StrandTwister has been tested using 100 β sheets simulated to 10 Å resolution and 39 β sheets that were computationally detected from cryo-EM maps at 4.4 to 7.4 Å resolutions. We find that StrandTwister can detect the traces of β strands fairly well for many β sheets with three or more β strands. The detection of β traces is limited by the identification of β sheet. This means that two-stranded β sheets and those β sheets in low-quality regions of a density map remain challenging. We discuss the results and challenges in β strand detection using three cases: gp10 of bacteriophage epsilon15 map (7.3 Å resolution), GroEL density (4.2 Å resolution), and E2 of Venezuelan equine encephalitis virus map (4.4 Å resolution).

RESULTS

StrandTwister has two major components. The first simplifies the voxels of a β sheet into a polynomial surface. The second, also the major step, identifies right-handed β twist from the polynomial surface model. The accuracy of detection was evaluated from two aspects: the two-way distance between the set of detected and that of observed β traces (Supplemental Experimental Procedures available online) and the percentage of the detected Cα atoms of observed β strands. An amino acid of a β strand is considered detected if the projection distance from its Cα atom to the corresponding detected β trace is less than 2.5 Å, which is about half of β strand spacing.

Right-Handed β Twist and β Strand Detection

Right-handed twist of a β sheet was first described by Cyrus Chothia in 1973 (Chothia, 1973). Salemme et al. suggested that the spatial configuration of a β sheets is isotropically stressed surface (Salemme, 1981). To understand the right-handed twist, one may thread the right-hand fingers along the β strands (Figure 3D). The natural curvature of the right hand would lift up the index finger and lower the pinky (Figure 3D).

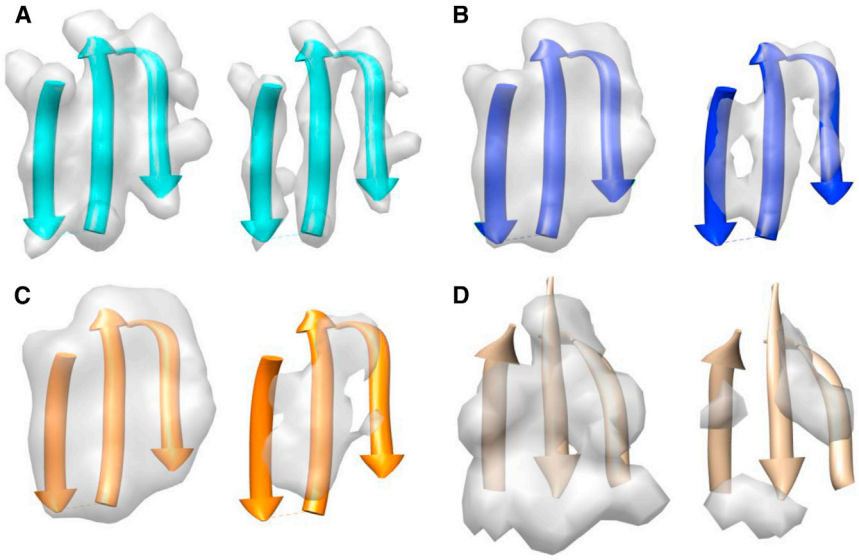

Figure 2. Density of a β Sheet at Different Resolutions

(A–C) The density was simulated using atomic structure of 1A12 (PDB accession number) and EMAN (Ludtke et al., 1999) at 6 Å in (A), 8 Å in (B), and 9 Å in (C).

(D) A β sheet was detected from an experimentally derived cryo-EM density map (EMD_5030) at 6.4 Å using SSETracer (Si and He, 2013). All data were visualized in Chimera (Pettersen et al., 2004).

The surface representation of density is shown at a lower (left) and a higher (right) threshold in (A) through (D).

The key idea in our method is the observation that the density of a β sheet reveals the right-handed twist very well (Figures 3A–3C), particularly at medium resolutions. In addition, we discovered that wrongly positioned β strands in the β sheet density can produce left-handed twist. For example, if one attempts to thread fingers into the paper (a wrong orientation) and fit the β sheet density in Figure 3A, it must be the left hand, not the right hand. This suggests that the correct orientation of the β strands can be identified by measuring the handedness and twist angles.

Performance on the Simulated Maps

The purpose of this test is to investigate if β strands can be traced from the density maps simulated at 10 Å resolution, at which the separation of β strands is not visible. StrandTwister was tested using 100 simulated β sheet density maps. For each β sheet density, possible sets of β traces were produced and sorted by the average main twist (AMT). Figure 4 shows the best of the top ten sets of detected β traces (red lines) for three cases with three, six, and nine β strands, respectively. In the case of sheet B of Protein Data Bank (PDB) structure 1T8H, one of the top ten detected sets appears to align with the β strands very well (Figure 4B). In this case, all six β traces were detected with a small two-way distance of 0.70 Å (Table 1, row 3 of six stranded). We observed that the top ten detected sets always include a set with close orientations to the actual β strand orientations. The detection is slightly better for smaller β sheets such as those with three to seven strands, if the number of detected β traces and the two-way distance are considered. However, some large β sheets were still well detected, such as the nine-stranded sheet 1UD9_B (Figure 4C). The two-way distance is only 1.07 Å in this case (Table 1), and most of the β traces are accurately detected. The error appears to be at the edge of the β sheets.

The test of 100 simulated β sheet density maps shows that one of the top ten ranked sets of β traces aligns very well with the observed set of β traces, with an overall two-way distance of 1.24 Å (Table 1, last two rows) for the detected β traces. To analyze the sensitivity of the detection, we measured the number of amino acids that were missed in the detection (see definition early in Results). For example, 1ATZ_A has five of the six β strands detected (Table 1, row 1 of six stranded), and four amino acids were missed. For the five β traces detected, the two-way distance is 1.18 Å. Among the 100 test cases, StrandTwister appears to be able to detect 81.48% of the β strands in one of the top ten ranked sets of β traces (Table 1).

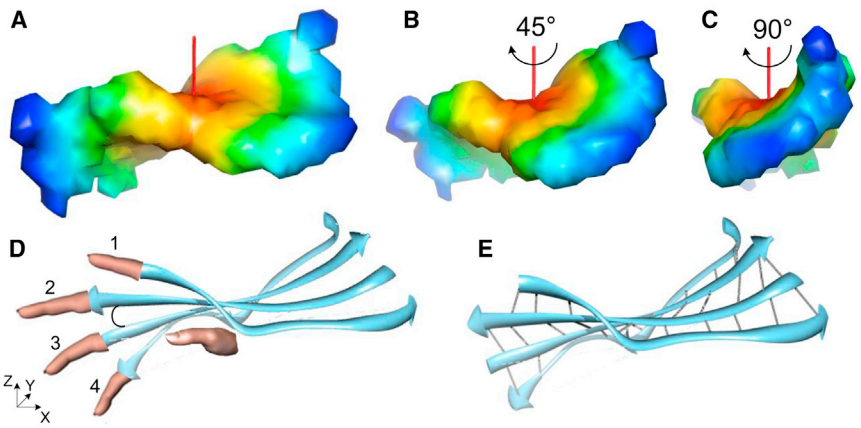

Figure 3. The Right-Handed Twist of a β Sheet

(A–C) The β sheet density was detected using SSETracer from cryo-EM map (EMD_1237) and is shown in 0° (A), 45° (B), and 90° (C) view around the normal vector (red line).

(D) The right-handed twist of β sheet 2GSY_sheet G. The strands are labeled such that the index finger (1) is at forefront and the pinky (4) is farthest from the page.

(E) The correspondent PDB structure of (A) is shown with hydrogen bonds (thin black lines).

See also Figure S1.

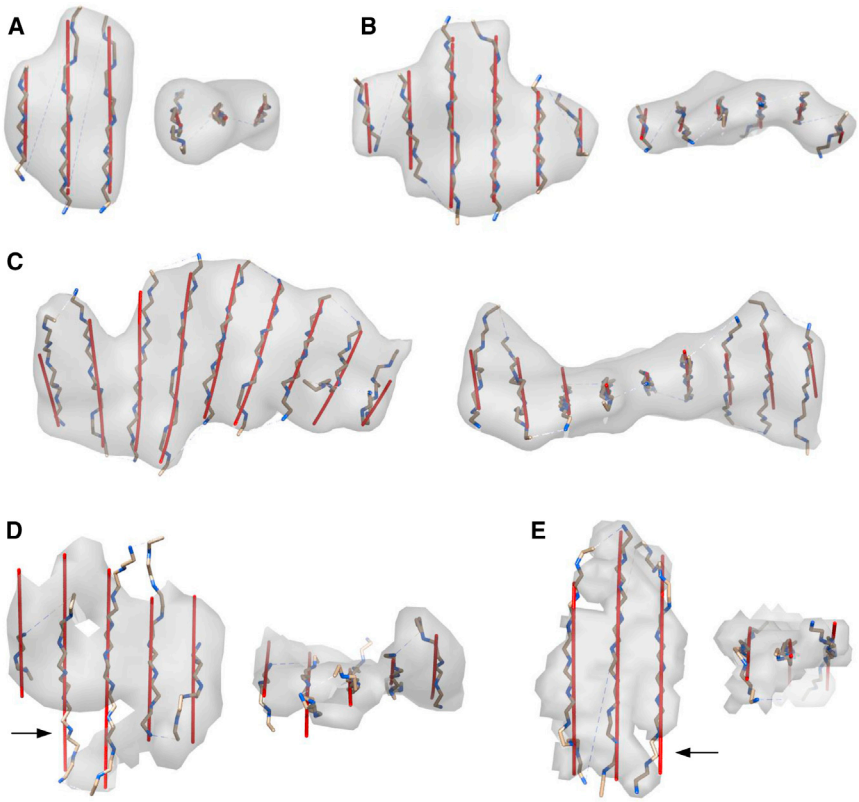

Figure 4. β Strand Detection from Simulated Density Maps at 10 Å and Experimental Cryo-EM Maps

The best of the top ten sets of detected β traces (red lines) are superimposed with the backbone of the β strands and the density maps (gray) for β sheets 1A4I_B in (A), 1T8H_B in (B), 1UD9_B in (C), EMD_2165_4B4T_1C in (D), and EMD_1780_3IZ5_Z in (E). The top view (left) and the side view (right) are shown in each case. The density (gray) of β sheet in (D) and (E) was detected using SSETracer. See also Figure S6.

β Strand Detection from Cryo-EM Maps

The performance of StrandTwister on the error-free simulated β sheet density shows the potential of our β strand detection using the principle of β sheet twist. In this section, we examine the performance of β strand detection using 39 β sheets, a large data set from experimentally derived cryo-EM maps. The results in this section represent the performance of both SSETracer and StrandTwister, because the error at either step will affect the results in Table 2.

Figure 4D shows the density detected by SSETracer from a cryo-EM map at 7.4 Å resolution. At this resolution, the separation of β strands is not visible, and some regions may have weak or missing density (arrows) due to errors in the data or in the identification of β sheet. However, the mathematical surface fitting appears to compensate for the missing density to some extent, because the surface is based on the overall density distribution of the detected β sheet. StrandTwister was able to detect all five strands in 2165_4B4T_1C (Figure 4D), and they align fairly well with the observed β traces. In this case, the two-way distance for the five strands is only 1.60 Å, and it detected 24 of 31 amino acids on the β sheet (Table 2).

The number of strands was correctly detected in 28 of the 39 cases (shown in Table 2 for the best of the top ten detections), in spite of the challenge from missing or extra density in the experimentally derived data. This is notable, because StrandTwister does not require the knowledge of the number of strands in β sheet during detection. Two different sampling sets of β traces may differ in the number of strands that are determined by the width of sheet perpendicular to the strand orientation. However, right-handed twist correctly distinguished between a three-stranded β sheet (with longer strands) and a six-stranded β sheet (with shorter strands) in the case of 1780_3IZ5_Z (Figure 4E), because sampling along the orientation perpendicular to the true orientation may result in six shorter β strands with left-handed twist. This suggests that the number of β strands and the positions of the β traces are intrinsic characters of β sheet density and are reflected in the β twist.

The results in Table 2 suggest that it is possible to detect traces of β strands from many β sheets identified from cryo-EM maps at medium resolutions. The overall two-way distance for the 39 test cases is only 1.66 Å for the best detected β traces (Table 2). StrandTwister was able to detect 81.87% of the β strand amino acids overall. Table 2 shows that once a β sheet region is identified roughly correctly, traces of β strands can be detected fairly well. All of the sheets in Table 2 have three or more β strands, because the identification of two-stranded sheet is not reliable. In the case of EMD_1237_2GSY (7.2 Å resolution), there is one unique chain (chain A) that contains six β sheets. SSETracer identified five β sheets (Table 2) but missed a two-stranded β sheet (sheet D) that contains four amino acids. Two extra short strands were wrongly detected by StrandTwister because of inaccurate boundary of identified β sheets. Note that the annotation of sheet F and sheet G in the PDB file corresponds to similar region, and hence we counted them as one sheet. EMD_1740_3C92 (6.8 Å resolution) has two unique chains (chains A and H) that contain five sheets. Four of the five sheets were detected well, although a two-stranded sheet (sheet P, with seven amino acids) was missed.

Detection of Four β Strands from gp10 Protein of Epsilon15

The backbone Cα trace of proteins in bacteriophage epsilon15 was derived from the cryo-EM density map (EMD_5678) at 4.5 Å resolution (Baker et al., 2013). Its staple protein, gp10, appears to contain two β sheets (dashed outlines in Figure 5D). The larger β sheet was predicted to contain four strands (Baker et al., 2013). In order to see if StrandTwister is able to detect the β strands from density maps at medium resolution, we applied StrandTwister to gp10 density that was extracted from epsilon15 map (EMD_1557) (Zhang et al., 2009) at 7.3 Å resolution.

Table 1. β Trace Detection from Simulated Density Maps at 10 Å Resolution

PDB ID[a]	Fit Error[b]	No. Detected[c]	Two-Way Distance[d]	No. Detected/ No. Observed AA[e]
Three Stranded				
1A12_B	1.29	3	0.95	16/17
1A12_D	1.25	3	0.93	15/17
1A4I_B	1.30	3	0.64	18/18
1A8D_B	1.18	3	0.75	12/12
1ATG_B	1.39	3	1.07	10/12
1AZO_A	1.52	3	0.82	16/18
1AZO_B	1.52	3	0.84	13/13
1B3A_A	1.45	3	0.82	14/15
1B5E_A	1.22	3	1.01	13/13
1BM8_A	1.81	3	1.11	16/18
1BTE_B	1.43	3	1.41	22/24
1BUP_C	1.54	3	1.45	16/19
1E0M_1	1.61	3	1.08	15/18
Four Stranded				
1A12_A	1.37	4	0.97	24/25
1A12_C	1.28	4	1.07	18/20
1A12_J	1.33	4	0.93	18/20
1A4I_C	1.20	4	0.76	18/18
1A8D_E	1.52	4	1.03	25/27
1AOP_1	1.56	4	1.08	18/21
1AQZ_D	1.59	3	1.29	17/23
1BUP_D	1.57	4	0.93	28/32
1C1D_C1	1.25	4	0.74	27/28
1CC8_A	1.50	4	0.98	25/27
1CCW_B	1.43	4	1.36	17/22
1DD9_A	1.40	3	1.74	15/21
1DS1_C	1.29	3	0.95	13/17
1Q38_1	1.43	4	0.94	23/27
1S04_1	1.56	4	1.37	14/16
2P8Y_1	1.37	4	0.98	19/20
2VZ1_AE	1.46	4	0.78	17/18
Five Stranded				
1AKY_A	1.44	5	1.22	22/25
1AOP_1	1.54	5	1.03	26/29
1AOP_2	1.75	5	1.78	21/38
1AOP_3	1.63	5	1.17	26/31
1B5E_D	1.40	5	1.88	18/29
1BUP_A	1.51	5	1.12	26/27
1C1D_A	1.57	4	1.19	29/37
1C7K_A	1.66	5	1.59	24/30
1CXQ_A	1.53	5	0.95	24/27
1DGW_C	1.46	4	1.00	27/32
1DMH_C	1.46	5	0.77	26/27
1DTD_B	1.53	4	1.10	23/28
1E2K_A	2.59	4	2.47	13/28
1MOL_S1	2.44	5	1.91	34/50
1ZH2_A	1.34	5	1.19	20/24
2JKX_AD	1.23	4	0.74	19/23
2VZ1_AD	1.23	4	0.66	19/23
Six Stranded				
1ATZ_A	1.51	5	1.18	36/40
1RV9_B	1.28	6	0.95	27/29
1T8H_B	1.13	6	0.70	28/28
1VLY_A	1.47	6	0.99	25/29
1YT3_A	1.20	6	0.90	29/31
2HKE_A	1.32	6	0.80	31/32
2P51_A	1.36	6	0.98	28/30
2QTR_A	1.56	5	1.85	22/31
2VBF_BA	1.51	6	0.96	29/31
2VOA_AB	1.38	6	1.00	29/32
2ZSG_A	1.47	6	0.82	28/30
3BL9_D	2.33	5	1.84	21/29
Seven Stranded				
1CHD_SH1	1.47	7	1.42	34/39
1D5T_D	1.78	6	1.48	47/56
1ELU_B	1.57	6	1.84	27/36
1FX2_A	1.87	7	0.95	40/45
1FYE_A	1.36	7	1.00	28/31
1G8K_B	1.43	7	1.06	28/29
2A6Z_A	1.49	7	0.94	42/47
2APJ_A	1.56	6	1.58	22/37
2DKJ_C	1.58	7	1.29	29/32
3BA1_B	1.41	7	0.92	25/26
Eight Stranded				
1DTD_A	1.68	8	1.15	40/47
1H2W_AJ	2.36	8	1.12	44/51
1HDO_AA	1.73	7	0.92	34/37
1JL0_A	1.56	8	0.85	48/52
1JOV_D	1.64	7	1.32	52/60
1JUH_A	1.73	8	1.04	40/44
1JW9_A	1.63	8	1.76	33/45
1KMV_B	1.96	7	1.65	34/49
1LAM_B	1.45	8	1.04	50/56
1M15_A	1.70	7	1.76	37/46
1M4L_A	1.80	8	0.99	41/48
1NKG_C	1.64	7	1.11	45/54
1ZLI_1	1.82	8	1.16	41/47
3RL6_1	1.73	7	1.89	35/51
8DFR_S1A	1.88	8	1.46	42/56
Nine Stranded				
1QNA_C	1.62	7	1.90	39/56
1UD9_B	2.02	9	1.07	49/58
1UWC_BA	1.68	8	1.35	41/53
2ABS_A	1.47	10	1.29	39/49

(*Continued on next page*)

Table 1. *Continued*

PDB ID[a]	Fit Error[b]	No. Detected[c]	Two-Way Distance[d]	No. Detected/ No. Observed AA[e]
2EAB_A	2.04	8	1.18	59/75
2VVG_AB	1.92	7	1.67	43/67
3DB7_A	2.21	9	1.90	37/51
3EN0_A	1.46	8	1.66	32/39
3FCX_A	2.19	9	1.71	48/66
3H9M_C	2.10	7	2.38	30/53
3HID_A	1.94	6	1.66	25/40
3HO6_A	2.03	9	1.89	38/53
Ten Stranded				
1IG0_B	1.81	9	2.13	33/55
1PE9_A	1.34	9	1.68	40/52
1V7W_A	1.77	9	1.26	59/70
2B0T_A	2.17	8	2.47	22/44
Average	1.59		1.24	2,834/3,478 = 81.48%
SD	0.29		0.42	

[a]The PDB accession number and sheet ID.
[b]The root-mean-square error (rmse, in Å) for the polynomial surface fitting. See also the description in the Supplemental Information.
[c]The number of detected β traces.
[d]The two-way distance (in Å) for the best of the top ten sets of detected β traces.
[e]The number of detected/the total number of amino acids in the β sheet.

At 4.5 Å resolution, the two sheets are shown as separate sheets (Figure 5D), and the separation of β strands is visible. However, this is not true for the 7.3 Å resolution map (Figure 5A). The lower sheet region has weak or missing density, and SSETracer detected only the upper β sheet (Figures 5A and 5B). The detected β sheet voxels appear to show the twist nature of that β sheet (Figure 5B). StrandTwister successfully detected four β traces from the detected β sheet of gp10 (red lines in Figure 5C). Our analysis of β twist in the 7.3 Å resolution map further supports the finding of four-stranded sheet in gp10. Note that only the Cα trace of the backbone is available in the PDB file, and therefore the β sheet is not defined. However, the Cα chain appears to resemble a four-stranded β sheet (Baker et al., 2013).

Detection of β Strands from GroEL Density Map EMD_5001

The quality of a density map may vary from region to region, so it is possible that not all β strands are well detected in a map at 4 to 5 Å resolution. The cryo-EM density map of GroEL (EMD_5001) was obtained at 4.2 Å resolution, from which the Cα trace (PDB 3CAU) was derived using de novo modeling (Ludtke et al., 2008). There are three other GroEL structures (1SS8, 1OEL, and 1XCK) that have been solved by X-ray crystallography (Bartolucci et al., 2005; Braig et al., 1995; Chaudhry et al., 2004). Although these four structures are slightly different, they all appear to have seven β sheets at approximately the same locations. We aligned the Cα trace of 3CAU with the three crystal structures using Matchmaker in Chimera. The main difference among the four structures appears to be at the upper domain, which contains sheets C, D, and E (Figures 6A and 6B).

SSETracer detected five of the seven sheets from the density monomer of EMD_5001 (Figure 6B). Two two-stranded β sheets (F and G) were missed because a two-stranded sheet can be confused with a helix (Figure 6A, orange arrows). Because the Cα trace of 3CAU does not have an annotation of secondary structures, we used the annotation in the other three X-ray structures to estimate the number of strands in the β sheets. However, the beginning and ending positions of β strands vary among the three structures. StrandTwister detected 14 of 19 traces of β strands from the density monomer of GroEL (Figure 6; Table S2 and Figure S4). The detected β traces for β sheet A and B appear to agree well with those in the four structures (Figures 6C and 6D). In fact, the two-way distances are only 1.28 and 1.13 Å, respectively, for 1SS8_A and 1SS8_B (Table S2). In terms of sheets C, D, and E, the annotation of β sheets is different among the X-ray structures. Figure S4 shows the set of β traces that best align with strands in 1OEL. Note that the detected β traces in Figure S4 agrees well with that of the X-ray structures (1SS8, 1OEL, and 1XCK), although the beginning and ending positions may differ. When the start and end of a strand are not accurately detected, two-way distance will be affected. For example, the two-way distance for 1SS8_C is 2.46 Å (Table S2), which is larger than that of 1SS8_B (1.13 Å). This is mostly because the detected traces are much longer than the observed ones (arrows in Figure S4A). Note that the detected β traces do not align with the Cα trace of 3CAU at sheet D (row 2 in Figure S4B). In fact, none of the top ten detected sets align well with 3CAU at sheet D.

Detection of β Strands from E2 of Venezuelan Equine Encephalitis Virus Density Map EMD_5276

The structure of Venezuelan equine encephalitis virus was resolved from the 4.4 Å resolution cryo-EM density map. E2 contains a transmembrane helix and 30 β strands that are on ten β sheets (Zhang et al., 2011) (see Table S2 for details of annotation). We isolated the density monomer of E2, which aligned with chain B of 3J0C. SSETracer detected five larger β sheets (N, K, O, T, and R) (Figure 7). Three two-stranded β sheets and two three-stranded β sheets were missed. Sheet Q (three stranded) is mostly a two-stranded twist and appears as a helix in the density. Sheet S (three stranded) is located at the outer surface of E2 (Figures 7A and 7B; Figure S5A) where the density is weak and has no obvious sheet property. We applied the Gaussian filter (Figure S5) to the outermost domain to enhance the weak density and were able to detect sheet R (Figure S5). StrandTwister detected 17 of 30 β traces in E2, suggesting that the majority of the β strands can be detected for β sheets with 3 or more β strands. The detected β traces align well with the corresponding observed β strands on sheets K, O, and T, with 1.47, 1.60, and 1.13 Å two-way distances, respectively (Table S2).

Cα Models Derived from β Traces

A rapid method has been developed to construct the backbone from the density of a β sheet for situations in which the resolution of the density map is high enough to resolve the density pattern of a β strand (i.e., ~3.8Å resolution) (Terwilliger, 2010). At the medium resolutions, however, multiple possible Cα models may be derived from a β trace. StrandTwister produces a possible Cα model from a set of β traces (Experimental

Table 2. Accuracy of β Strand Detection for the Experimentally Derived Cryo-EM Maps

EMDB_PDB ID[a]	Resolution[b]	Fit Error[c]	No. Detected/ No. Observed Strands[d]	Two-Way Distance[e]	No. Detected/ No. Observed AA[f]
1237_2GSY_A	7.2	2.19	4/4	2.34	13/22
1237_2GSY_B		2.30	5/5	1.90	26/28
1237_2GSY_C		2.37	5/6	1.90	36/41
1237_2GSY_E		1.80	5/4	3.03	23/47
1237_2GSY_G		2.16	5/4	1.85	36/48
1740_3C92_A	6.8	1.29	5/5	1.71	24/27
1740_3C92_B		1.62	5/5	1.62	29/31
1740_3C92_O		1.46	4/5	1.04	24/28
1740_3C92_Q		1.47	5/5	1.59	26/27
1780_3IZ5_AC	5.5	2.15	4/4	1.58	21/25
1780_3IZ5_AH		1.33	4/4	1.60	14/15
1780_3IZ5_AI		1.60	3/3	1.56	18/22
1780_3IZ5_AS		1.44	4/3	1.72	12/19
1780_3IZ5_AT		1.35	4/4	1.53	19/20
1780_3IZ5_AY		1.61	4/5	1.96	12/16
1780_3IZ5_F		1.45	5/5	1.41	27/30
1780_3IZ5_H		1.28	3/3	1.32	15/17
1780_3IZ5_I		2.14	3/4	1.37	17/22
1780_3IZ5_J		1.35	3/3	1.46	12/14
1780_3IZ5_K		1.31	4/4	1.63	16/19
1780_3IZ5_L		1.81	4/5	2.17	19/21
1780_3IZ5_R		1.75	4/4	1.73	23/32
1780_3IZ5_W		2.00	4/4	1.93	19/23
1780_3IZ5_Z		1.65	3/3	1.00	27/28
1780_3IZ6_AF		1.51	3/3	0.89	14/14
1780_3IZ6_D		1.64	3/3	1.58	12/13
1780_3IZ6_F		2.01	4/4	1.95	15/20
1780_3IZ6_I		1.51	4/5	1.57	15/20
1829_2WWL_CA	5.6	1.46	3/4	1.72	15/21
1829_2WWL_CB		1.59	4/4	1.91	22/24
1829_2WWQ_SA		2.18	3/3	1.80	21/26
1829_2WWQ_TA		1.11	3/3	1.03	13/13
2165_4B4T_1A	7.4	1.39	4/5	2.18	19/31
2165_4B4T_1C		1.47	5/5	1.60	24/31
2165_4B4T_AA		1.28	5/5	1.38	23/26
5036_3FIH_P	6.7	1.02	3/3	1.94	13/17
5276_3J0C_K	4.4	1.16	4/4	1.47	20/22
5276_3J0C_O		1.52	3/3	1.60	12/15
5276_3J0C_T		1.57	3/3	1.13	17/17
Average		1.62		1.66	763/932 = 81.87%
SD		0.35		0.40	

[a]The EMDB and PDB accession numbers and sheet ID.
[b]Resolution of the density map.
[c]The rmse (in Å) for the polynomial surface fitting. See also the description in the Supplemental Information.
[d]The number of β traces in the best of the top ten detected sets/the number of β strands in the true structure.
[e]The two-way distance (in Å) between the observed β traces and the detected β traces for the best of the top ten detected sets.
[f]The number of detected/the total number of amino acids in the β sheet.

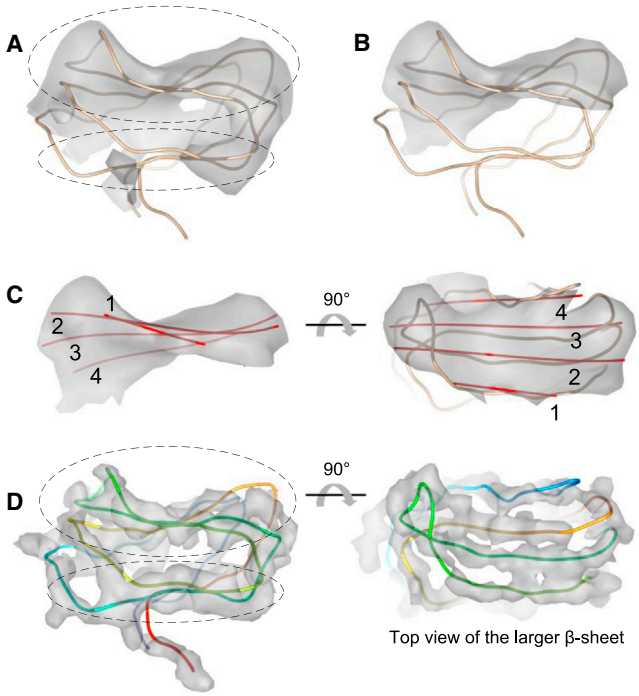

Figure 5. β Strand Detection from the 7.3 Å Resolution Map of Epsilon15

(A) The density region (gray surface) of gp10 protein was extracted from bacteriophage epsilon15 density map EMD_1557 at 7.3 Å resolution. The Cα chain of gp10 (chain I of PDB 3J40) was superimposed with the density.

(B) The β sheet density region detected using SSETracer.

(C) The β strands detected (red lines) using StrandTwister are superimposed on the density of β sheet (left: side view) and the Cα trace of β strands (right: top view).

(D) The superposition of gp10 density at 4.5 Å resolution (EMD_5678) and PDB structure (chain I of 3J40) is shown as a side view (left) and a top view (right, for the larger sheet region).

See also Figure S3 and Table S2.

Procedures; Figure S7). A test using 39 sets of β traces detected from cryo-EM density maps shows that the models have good overall accuracy of 2.56 Å root-mean-square deviation (rmsd) for 84% of the Cα atoms in the true β sheets (Table S3). This error is reasonable, because the input density maps have resolutions of about 4.4 to 7.4 Å.

DISCUSSION

As expected, accurate detection of β strands depends on accurate identification of a β sheet. The boundary of the identified β sheet may affect β strand detection. In most cases, an inaccurate boundary can result in a longer or shorter detected β strand, and such error is reflected in the two-way distance value. The missing density at the upper region of the β sheet in Figure 4D resulted in missing three amino acids in the detection. The extra density (box in Figure S6) affected the curvature and size of the β sheet, and the detected set was slightly off, with a two-way distance of 2.34 Å (first row of Table 2). Our implementation ignored all the short strands less than 6 Å in length. This may also be responsible for some of the missing strands.

StrandTwister was tested using 39 β sheets, and the results were analyzed in details for three case studies. The conclusion from the tests appears to be twofold. First, many β traces can be detected from density maps at medium resolutions. Our proposed idea to use β twist in detection appears to be effective once a β sheet region is identified approximately. One of the top ten sets of β traces contains a set with close estimation to the observed β traces, particularly at the central region of a β sheet. Fine adjustment is needed to improve the detection near the edge of β sheet, where special properties have been observed (Richardson and Richardson, 2002). Second, the detection error comes mainly from two situations, one being the inaccurate boundary of β sheets and the other being misidentified β sheets. Both are long-standing challenges for β sheet identification. The limitation of β strand detection arises mostly from the identification of a β sheet.

In spite of the errors in identifying short helices and two-stranded β sheets, major or larger helices and sheets can be detected from the medium-resolution maps. Our results in this paper add to the statement that major helices and those β strands on larger sheets can be traced. Our previous results and those from other groups have shown that the topology of major secondary structures may not rely on the detection of all secondary structures. In many cases, the topology of major helices is correctly predicted without the detection of short helices (Al Nasr et al., 2012; Lindert et al., 2012; Lu et al., 2008). Deriving atomic structures from density maps at medium resolutions will inevitably involve sophisticated modeling of uncertainties. The methodology in de novo modeling from the medium resolution density maps has been improved recently to work with a large number of helices (Al Nasr et al., 2014), to work with complicated skeletons (Al Nasr et al., 2013), and to build the atomic chains in modeling (Baker et al., 2012b; Lindert et al., 2012). However, the work has been tested mostly using the true position of β strands for density maps at medium resolutions. StrandTwister detects the traces of β strands for major β sheets. With the β traces, de novo modeling is expected to move a significant step ahead. The current topology determination method will be extended to both α traces and β traces. We have shown that additional constraints can be added to represent popular β strand pairing during topology determination (Al Nasr et al., 2014). We illustrate an effective method to build a Cα backbone using β traces (Supplemental Experimental Procedures). However, density errors in the cryo-EM maps at medium resolutions determine that multiple sets of possible β strands must be generated. The correct set must be identified when it is modeled together with other parts of the chain. Another potential impact of β traces lies in the representation of secondary structures in density maps. The relative locations of α traces have been used as signatures to search for atomic structures with similar folds (Lasker et al., 2007). Now it is possible to include β traces as well.

Deriving atomic structures from the medium resolution cryo-EM density maps is a challenging problem. Although a number of methods exist to detect α helices, the detection of β strands from medium-resolution cryo-EM maps has been an open problem since the first attempt in 2004 (Kong et al., 2004). We propose a method to detect both the number of β strands and the β traces directly from medium-resolution density data using

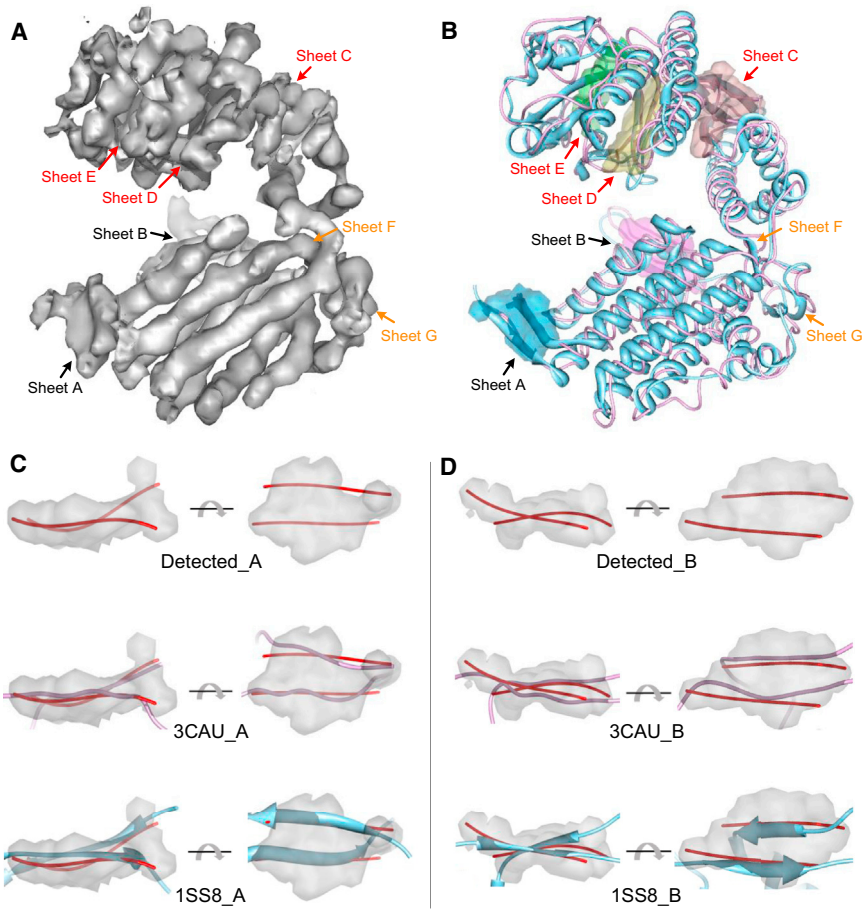

Figure 6. β Strand Detection from the Density Map of GroEL at 4.2 Å Resolution

(A) The monomer density (gray) of GroEL extracted from density map EMD_5001.

(B) Five β sheet density regions (colored density) identified using SSETracer are superimposed on chain A of PDB 3CAU (purple Cα trace) and chain A of PDB 1SS8 (cyan ribbon).

(C and D) The side view (left) and the top view (right) of the β-traces detected (red lines) using StrandTwister are superimposed with PDB structures of 3CAU (purple Cα trace) and 1SS8 (cyan ribbon) for sheet A in (C) and sheet B in (D).

See also Figure S4 and Table S2.

the intrinsic twist of a β sheet. To our knowledge, this is the second attempt to address the problem of β strand detection in 10 years, and we provide an optimistic answer to this problem using a completely different approach.

EXPERIMENTAL PROCEDURES

Polynomial Fitting of β Sheet Density

A β sheet generally appears as a thin layer of density at medium resolutions. In order to capture the overall shape of β sheet density and to make strand detection less sensitive to errors in density data, a polynomial surface (Equation 1) was first determined by fitting the density voxels (Supplemental Experimental Procedures and Figure S1). Here, (x,y,z) is the coordinate of a voxel point (Si and He, 2013).

$$z = Ax^3 + By^3 + Cx^2 + Dy^2 + Ex^2y + Fy^2x + Gxy + Hx + Iy + J \quad \text{(Equation 1)}$$

Right-Handed β Twist and β Strand Detection

The points on the polynomial surface (yellow in Figures S1B and S1C) were used to measure handedness and twist angles. The handedness and twist angles were calculated for each set of β traces. Each set of β traces was generated on the basis of the observation that β strands are roughly parallel to each other, with two neighboring strands forming a small twist angle (Supplemental Experimental Procedures and Figure S1).

Let V_1 and V_2 represent two neighboring β traces, and let V_{12} represent the vector pointing from line segment V_1 to V_2 and having the shortest distance between them. A right-handed twist requires the following:

- $(V_1 \times V_2)\cdot V_{12} > 0$, if V_1 and V_2 are on the antiparallel β strands
- $(V_1 \times V_2)\cdot V_{12} < 0$, if V_1 and V_2 are on the parallel β strands.

In principle, $V_1 \times V_2$ is on the line of V_{12}, either having the same direction as V_{12} in which $((V_1 \times V_2)\cdot V_{12}) > 0$, or having the opposite direction as V_{12} in which $((V_1 \times V_2)\cdot V_{12}) < 0$. Suppose that V_1 and V_2 are on antiparallel β strands, as in the first case; a right-handed twist will have $V_1 \times V_2$ in the same direction as V_{12} (Figure S1E). A left-handed twist will result in $((V_1 \times V_2)\cdot V_{12}) < 0$. Because a twist angle is often small, we use the acute angle formed by V_1 and V_2 as the interstrand twist angle (Figure S1E). A similar principle applies in the second case, in which V_1 and V_2 are on the parallel β strands.

Because there is a handedness and a twist angle for each pair of neighboring β traces, the overall handedness and twist angle of the entire set of β traces need to be determined. Those sets having right-handed twist for all neighboring strands were first selected. Such sets were further evaluated by the AMT angle to select the best one (Figure S2). AMT is the average of three consecutive interstrand twist angles: the largest interstrand twist angle and that to its left and to its right, respectively. The assumption is that the correct set of β traces is expected to have near maximum overall twist (Supplemental Experimental Procedures and Table S1). The top ten (first ten) sets of β traces with the largest AMT were detected as potential sets.

Simulated and Cryo-EM Data Collection

The simulated density maps were derived from PDB structures that were randomly collected with the following requirements: (1) the β sheets are regular sheets rather than barrels, and (2) the number of strands is between three and ten. The atomic structures of β sheets were used to generate β sheet density maps at 10 Å resolution using EMAN, popular software to produce simulated density (Ludtke et al., 1999), with a step size of 1 Å/pixel. A density threshold of 75% (~4–6 SDs) was set to the isosurface value.

Seven cryo-EM maps (Table 2) were collected from EMDB with resolutions between 4.4 and 7.4 Å. All seven density maps were aligned with their corresponding PDB structures at download except EMD_1237, for which we aligned manually with the help of the Fit in Map function of UCSF Chimera (Pettersen et al., 2004). The density region belonging to a chain of a protein was isolated from the molecular complex, such as a virus, using the PDB structure as an envelope. Once the density of the entire chain of a protein was obtained, SSETracer (Si and He, 2013) was used to identify β sheets from it. We recently modified SSELearner (Si et al., 2012) into a simple tool, SSETracer, for the identification of both helices and β sheets. Such identified β sheet density voxels (shown as a blue surface view in Figure 1B) were then forwarded to StrandTwister.

To extract the density of gp10, we first manually fit the chains of PDB structure 3J40 into the 7.3 Å map (EMD_1557) and refined with the Fit in Map option (Figure S3) in Chimera. The density region of gp10 (gray in Figure 5A) was manually extracted using the guidance of the fitted gp10 structure (see

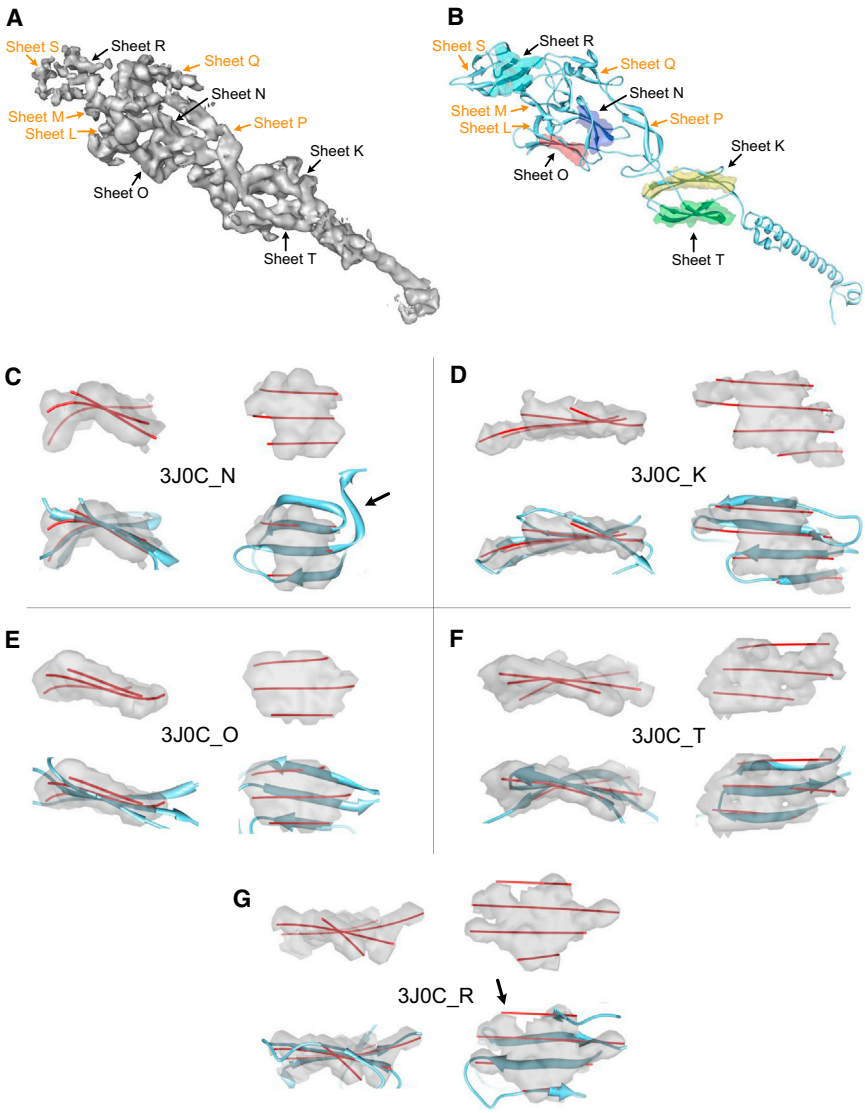

Figure 7. β Strand Detection from the Density Map of E2 in Encephalitis Virus: EMD_5276

(A) E2 monomer density (gray) at 4.4 Å resolution.

(B) β Sheet density regions (colored density) identified using SSETracer are superimposed on chain B of PDB 3J0C (cyan ribbon).

(C–G) The side view (left) and the top view (right) of detected β traces (red lines) using StrandTwister are superimposed with the observed β strands of sheets N in (C), K in (D), O in (E), T in (F), and R in (G). The density shown in (G) was obtained after applying Gaussian filter to enhance density connectivity at the outer domain.

See also Figure S5 and Table S2.

Figure S3). The envelope of gp10 protein can be distinguished at 7.3 Å resolution (box in Figure S3B).

Construction of the Cα Model from β Traces

We investigated a method to construct the Cα model by enforcing both β traces and the general rules observed from true structures of β sheets (Supplemental Experimental Procedures), so that the model is along the β traces and appears as β strands. The Cα modeling program takes a set of β traces and the density of the isolated β sheet as input. It output a Cα model for the β sheet in PDB format. The accuracy of the Cα models built for 39 sets of β traces detected from cryo-EM maps was based on two aspects: the number of the matched Cα atoms out of the total number of atoms in the observed β sheet and the rmsd for the matched Cα atoms.

Implementation

StrandTwister was written in C++. The executable software is downloadable, and the results in both simulated maps and experimental cryo-EM maps are viewable online at http://www.cs.odu.edu/~jhe/software/strandtwister. The input density map of a β sheet is in MRC format. The output of a set of β traces is in PDB format viewable in UCSF Chimera. StrandTwister is a fully automatic tool. The only parameter is the density threshold, a parameter for estimating the size of the β sheet density.

All tests were performed on a Linux machine with a 4x Intel Xeon X7550 at 2.00 GHz, with 128 GB of random-access memory. The execution time depends on the size of individual input MRC file of β sheet density. For example, the program takes 90.7 s to generate all outputs for an input β sheet density map with a box size 123 × 145 × 192 and a file size of 13 MB.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.str.2014.08.017.

AUTHOR CONTRIBUTIONS

Both authors participated in the development of the method and the manuscript. J.H. is involved in project design. D.S. is involved in implementation.

ACKNOWLEDGMENTS

We would like to thank Dr. Shuiwang Ji for helpful discussion and Nalin Ranjan for assistance in editing the manuscript. The work in this paper is partially

supported by the startup fund, Multidisciplinary Seed Funding and Modeling & Simulation Fellowship of Old Dominion University, and NSF_Bio_1356621.

Received: February 4, 2014
Revised: August 7, 2014
Accepted: August 8, 2014
Published: October 9, 2014

REFERENCES

Al Nasr, K., Sun, W., and He, J. (2010). Structure prediction for the helical skeletons detected from the low resolution protein density map. BMC Bioinformatics *11* (*Suppl 1*), S44.

Al Nasr, K., Ranjan, D., Zubair, M., and He, J. (2011). Ranking valid topologies of the secondary structure elements using a constraint graph. J. Bioinform. Comput. Biol. *9*, 415–430.

Al Nasr, K., Chen, L., Si, D., Ranjan, D., Zubair, M., and He, J. (2012). Building the initial chain of the proteins through de novo modeling of the cryo-electron microscopy volume data at the medium resolutions. Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine, 490–497.

Al Nasr, K., Liu, C., Rwebangira, M., Burge, L., and He, J. (2013). Intensity-based skeletonization of CryoEM gray-scale images using a true segmentation-free algorithm. IEEE/ACM Trans. Comput. Biol. Bioinformatics *10*, 1289–1298.

Al Nasr, K., Ranjan, D., Zubair, M., Chen, L., and He, J. (2014). Solving the secondary structure matching problem in cryo-EM de novo modeling using a constrained k-shortest path graph algorithm. IEEE/ACM Trans. Comput. Biol. Bioinformatics *11*, 419–430.

Anger, A.M., Armache, J.P., Berninghausen, O., Habeck, M., Subklewe, M., Wilson, D.N., and Beckmann, R. (2013). Structures of the human and Drosophila 80S ribosome. Nature *497*, 80–85.

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics *22*, 195–201.

Baker, M.L., Ju, T., and Chiu, W. (2007). Identification of secondary structure elements in intermediate-resolution density maps. Structure *15*, 7–19.

Baker, M.L., Zhang, J., Ludtke, S.J., and Chiu, W. (2010). Cryo-EM of macromolecular assemblies at near-atomic resolution. Nat. Protoc. *5*, 1697–1708.

Baker, M.L., Abeysinghe, S.S., Schuh, S., Coleman, R.A., Abrams, A., Marsh, M.P., Hryc, C.F., Ruths, T., Chiu, W., and Ju, T. (2011). Modeling protein structure at near atomic resolutions with Gorgon. J. Struct. Biol. *174*, 360–373.

Baker, M.L., Baker, M.R., Hryc, C.F., Ju, T., and Chiu, W. (2012a). Gorgon and pathwalking: macromolecular modeling tools for subnanometer resolution density maps. Biopolymers *97*, 655–668.

Baker, M.R., Rees, I., Ludtke, S.J., Chiu, W., and Baker, M.L. (2012b). Constructing and validating initial Cα models from subnanometer resolution density maps with pathwalking. Structure *20*, 450–463.

Baker, M.L., Hryc, C.F., Zhang, Q., Wu, W., Jakana, J., Haase-Pettingell, C., Afonine, P.V., Adams, P.D., King, J.A., Jiang, W., and Chiu, W. (2013). Validated near-atomic resolution structure of bacteriophage epsilon15 derived from cryo-EM and modeling. Proc. Natl. Acad. Sci. USA *110*, 12301–12306.

Bartolucci, C., Lamba, D., Grazulis, S., Manakova, E., and Heumann, H. (2005). Crystal structure of wild-type chaperonin GroEL. J. Mol. Biol. *354*, 940–951.

Beck, F., Unverdorben, P., Bohn, S., Schweitzer, A., Pfeifer, G., Sakata, E., Nickell, S., Plitzko, J.M., Villa, E., Baumeister, W., and Förster, F. (2012). Near-atomic resolution structural model of the yeast 26S proteasome. Proc. Natl. Acad. Sci. USA *109*, 14870–14875.

Biswas, A., Si, D., Al Nasr, K., Ranjan, D., Zubair, M., and He, J. (2012). Improved efficiency in cryo-EM secondary structure topology determination from inaccurate data. J. Bioinform. Comput. Biol. *10*, 1242006.

Braig, K., Adams, P.D., and Brünger, A.T. (1995). Conformational variability in the refined structure of the chaperonin GroEL at 2.8 A resolution. Nat. Struct. Biol. *2*, 1083–1094.

Chan, K.Y., Trabuco, L.G., Schreiner, E., and Schulten, K. (2012). Cryo-electron microscopy modeling by the molecular dynamics flexible fitting method. Biopolymers *97*, 678–686.

Chaudhry, C., Horwich, A.L., Brunger, A.T., and Adams, P.D. (2004). Exploring the structural dynamics of the E.coli chaperonin GroEL using translation-libration-screw crystallographic refinement of intermediate states. J. Mol. Biol. *342*, 229–245.

Chiu, W., Baker, M.L., Jiang, W., Dougherty, M., and Schmid, M.F. (2005). Electron cryomicroscopy of biological machines at subnanometer resolution. Structure *13*, 363–372.

Chothia, C. (1973). Conformation of twisted beta-pleated sheets in proteins. J. Mol. Biol. *75*, 295–302.

Crowther, R.A., Kiselev, N.A., Böttcher, B., Berriman, J.A., Borisova, G.P., Ose, V., and Pumpens, P. (1994). Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. Cell *77*, 943–950.

Del Palu, A., He, J., Pontelli, E., and Lu, Y. (2006). Identification of alpha-helices from low resolution protein density maps. Proceeding of Computational Systems Bioinformatics Conference (CSB), 89–98.

Esquivel-Rodríguez, J., and Kihara, D. (2013). Computational methods for constructing protein structure models from 3D electron microscopy maps. J. Struct. Biol. *184*, 93–102.

Hashem, Y., des Georges, A., Fu, J., Buss, S.N., Jossinet, F., Jobe, A., Zhang, Q., Liao, H.Y., Grassucci, R.A., Bajaj, C., et al. (2013). High-resolution cryo-electron microscopy structure of the Trypanosoma brucei ribosome. Nature *494*, 385–389.

Henderson, R., Sali, A., Baker, M.L., Carragher, B., Devkota, B., Downing, K.H., Egelman, E.H., Feng, Z., Frank, J., Grigorieff, N., et al. (2012). Outcome of the first electron microscopy validation task force meeting. Structure *20*, 205–214.

Hryc, C.F., Chen, D.H., and Chiu, W. (2011). Near-atomic-resolution cryo-EM for molecular virology. Curr. Opin. Virol. *1*, 110–117.

Jiang, W., Baker, M.L., Ludtke, S.J., and Chiu, W. (2001). Bridging the information gap: computational tools for intermediate resolution structure interpretation. J. Mol. Biol. *308*, 1033–1044.

Jiang, W., Baker, M.L., Jakana, J., Weigele, P.R., King, J., and Chiu, W. (2008). Backbone structure of the infectious epsilon15 virus capsid revealed by electron cryomicroscopy. Nature *451*, 1130–1134.

Ju, T., Baker, M.L., and Chiu, W. (2007). Computing a family of skeletons of volumetric models for shape description. Comput. Aided Des. *39*, 352–360.

Kong, Y., and Ma, J. (2003). A structural-informatics approach for mining beta-sheets: locating sheets in intermediate-resolution density maps. J. Mol. Biol. *332*, 399–413.

Kong, Y., Zhang, X., Baker, T.S., and Ma, J. (2004). A structural-informatics approach for tracing beta-sheets: building pseudo-C(alpha) traces for beta-strands in intermediate-resolution density maps. J. Mol. Biol. *339*, 117–130.

Lasker, K., Dror, O., Shatsky, M., Nussinov, R., and Wolfson, H.J. (2007). EMatch: discovery of high resolution structural homologues of protein domains in intermediate resolution cryo-EM maps. IEEE/ACM Trans. Comput. Biol. Bioinformatics *4*, 28–39.

Lasker, K., Förster, F., Bohn, S., Walzthoeni, T., Villa, E., Unverdorben, P., Beck, F., Aebersold, R., Sali, A., and Baumeister, W. (2012). Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. Proc. Natl. Acad. Sci. USA *109*, 1380–1387.

Lawson, C.L., Baker, M.L., Best, C., Bi, C., Dougherty, M., Feng, P., van Ginkel, G., Devkota, B., Lagerstedt, I., Ludtke, S.J., et al. (2011). EMDataBank.org: unified data resource for CryoEM. Nucleic Acids Res. *39*, D456–D464.

Lindert, S., Alexander, N., Wötzel, N., Karakaş, M., Stewart, P.L., and Meiler, J. (2012). EM-fold: de novo atomic-detail protein structure determination from medium-resolution density maps. Structure *20*, 464–478.

Lu, Y., He, J., and Strauss, C.E. (2008). Deriving topology and sequence alignment for the helix skeleton in low-resolution protein density maps. J. Bioinform. Comput. Biol. *6*, 183–201.

Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. J. Struct. Biol. *128*, 82–97.

Ludtke, S.J., Baker, M.L., Chen, D.H., Song, J.L., Chuang, D.T., and Chiu, W. (2008). De novo backbone trace of GroEL from single particle electron cryomicroscopy. Structure *16*, 441–448.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Richardson, J.S., and Richardson, D.C. (2002). Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. Proc. Natl. Acad. Sci. USA *99*, 2754–2759.

Rossmann, M.G. (2000). Fitting atomic models into electron-microscopy maps. Acta Crystallogr. D Biol. Crystallogr. *56*, 1341–1349.

Rusu, M., and Wriggers, W. (2012). Evolutionary bidirectional expansion for the tracing of alpha helices in cryo-electron microscopy reconstructions. J. Struct. Biol. *177*, 410–419.

Salemme, F.R. (1981). Conformational and geometrical properties of beta-sheets in proteins. III. Isotropically stressed configurations. J. Mol. Biol. *146*, 143–156.

Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. *234*, 779–815.

Schröder, G.F., Brunger, A.T., and Levitt, M. (2007). Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution. Structure *15*, 1630–1641.

Si, D., and He, J. (2013). Beta-sheet detection and representation from medium resolution cryo-EM density maps. Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics, 764–770.

Si, D., Ji, S., Nasr, K.A., and He, J. (2012). A machine learning approach for the identification of protein secondary structure elements from electron cryo-microscopy density maps. Biopolymers *97*, 698–708.

Sun, W., and He, J. (2009). Native secondary structure topology has near minimum contact energy among all possible geometrically constrained topologies. Proteins *77*, 159–173.

Terwilliger, T.C. (2010). Rapid model building of beta-sheets in electron-density maps. Acta Crystallogr. D Biol. Crystallogr. *66*, 276–284.

Topf, M., Baker, M.L., Marti-Renom, M.A., Chiu, W., and Sali, A. (2006). Refinement of protein structures by iterative comparative modeling and cryoEM density fitting. J. Mol. Biol. *357*, 1655–1668.

Topf, M., Lasker, K., Webb, B., Wolfson, H., Chiu, W., and Sali, A. (2008). Protein structure fitting and refinement guided by cryo-EM density. Structure *16*, 295–307.

Trabuco, L.G., Villa, E., Mitra, K., Frank, J., and Schulten, K. (2008). Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. Structure *16*, 673–683.

Yu, X., Jin, L., and Zhou, Z.H. (2008). 3.88 A structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy. Nature *453*, 415–419.

Yu, Z., and Bajaj, C. (2008). Computational approaches for automatic structural analysis of large biomolecular complexes. IEEE/ACM Trans. Comput. Biol. Bioinformatics *5*, 568–582.

Zhang, J., Nakamura, N., Shimizu, Y., Liang, N., Liu, X., Jakana, J., Marsh, M.P., Booth, C.R., Shinkawa, T., Nakata, M., and Chiu, W. (2009). JADAS: a customizable automated data acquisition system and its application to ice-embedded single particles. J. Struct. Biol. *165*, 1–9.

Zhang, J., Baker, M.L., Schröder, G.F., Douglas, N.R., Reissmann, S., Jakana, J., Dougherty, M., Fu, C.J., Levitt, M., Ludtke, S.J., et al. (2010). Mechanism of folding chamber closure in a group II chaperonin. Nature *463*, 379–383.

Zhang, R., Hryc, C.F., Cong, Y., Liu, X., Jakana, J., Gorchakov, R., Baker, M.L., Weaver, S.C., and Chiu, W. (2011). 4.4 Å cryo-EM structure of an enveloped alphavirus Venezuelan equine encephalitis virus. EMBO J. *30*, 3854–3863.

Zhang, X., Ge, P., Yu, X., Brannan, J.M., Bi, G., Zhang, Q., Schein, S., and Zhou, Z.H. (2013). Cryo-EM structure of the mature dengue virus at 3.5-Å resolution. Nat. Struct. Mol. Biol. *20*, 105–110.

Zhao, G., Perilla, J.R., Yufenyuy, E.L., Meng, X., Chen, B., Ning, J., Ahn, J., Gronenborn, A.M., Schulten, K., Aiken, C., and Zhang, P. (2013). Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. Nature *497*, 643–646.

Zhou, Z.H. (2008). Towards atomic resolution structural determination by single-particle cryo-electron microscopy. Curr. Opin. Struct. Biol. *18*, 218–228.

